Vol. 71, No. 5

tir- and *stx*-Positive *Escherichia coli* in Stream Waters in a Metropolitan Area

James A. Higgins,[1]* Kenneth T. Belt,[2] Jeffrey S. Karns,[1] Jonathan Russell-Anelli,[3] and Daniel R. Shelton[1]

USDA-ARS, Bldg. 173, 10300 Baltimore Blvd., Beltsville, Maryland[1]; NE USDA Forest Service, 134 TRC Bldg., University of Maryland at Baltimore County, Baltimore, Maryland[2]; and Department of Crop and Soil Sciences, Cornell University, Ithaca, New York[3]

Diarrheagenic *Escherichia coli*, which may include the enteropathogenic *E. coli* and the enterohemorrhagic *E. coli*, are a significant cause of diarrheal disease among infants and children in both developing and developed areas. Disease outbreaks related to freshwater exposure have been documented, but the presence of these organisms in the urban aquatic environment is not well characterized. From April 2002 through April 2004 we conducted weekly surveys of streams in the metropolitan Baltimore, Md., area for the prevalence of potentially pathogenic *E. coli* by using PCR assays targeting the *tir* and *stx*$_1$ and *stx*$_2$ genes. Coliforms testing positive for the presence of the *tir* gene were cultured from 653 of 1,218 samples (53%), with a greater prevalence associated with urban, polluted streams than in suburban and forested watershed streams. Polluted urban streams were also more likely to test positive for the presence of one of the *stx* genes. Sequence analysis of the *tir* amplicon, as well as the entire *tir* gene from three isolates, indicated that the pathogenic *E. coli* present in the stream waters has a high degree of sequence homology with the *E. coli* O157:H7 serotype. Our data indicate that pathogenic *E. coli* are continually deposited into a variety of stream habitats and suggest that this organism may be a permanent member of the gastrointestinal microflora of humans and animals in the metropolitan Baltimore area.

Enteropathogenic and enterohemorrhagic *Escherichia coli* (EPEC and EHEC, respectively) are important causes of severe diarrheal disease in the developing and developed worlds. The ability to cause attachment and effacing lesions among the intestinal epithelium is a hallmark of some groups of diarrheogenic *E. coli*; a variety of genes coding for proteins mediating the attachment and effacing phenotype are present on a 50- to 70-MDa plasmid (2, 6, 11, 26). A chromosomally located gene important in pathogenesis is *tir*, which codes for Tir, a 78-kDa protein serving as a receptor for the outer membrane protein intimin (2). Among different serotypes of diarrheogenic *E. coli*, the *tir* gene ranges in length from 538 (*E. coli* O26:H−) to 558 amino acids (*E. coli* O157:H7), with two putative transmembrane domains spanning amino acids 235 to 255 and 362 to 389 in *E. coli* O157:H7 (5). Another *E. coli* disease mechanism is mediated by Shiga-like toxins, which interfere with protein synthesis; the *stx*$_1$ and *stx*$_2$ genes are phage encoded and represent important molecular diagnostic targets for EHEC strains (25).

The pathobiology of EPEC and EHEC is the subject of intensive research; however, most of the existing information on the epidemiology of these organisms has been derived from clinical reports, and little is known regarding the prevalence and survival of these categories of *E. coli* in the environment. Traditional methods for monitoring water samples for the presence of bacteria, such as fecal coliform or enterococcus counts, are incapable of discriminating between nonpathogenic

and pathogenic *E. coli*, much less detecting EPEC or EHEC per se (10). We were interested in surveying stream water samples from watersheds representing a variety of land use and hydrologic scales for the presence of putative EHEC and EPEC strains via the use of PCR for the *tir* and *stx*$_1$ and *stx*$_2$ genes. It was our assumption that these strains would constitute a small percentage of the total *E. coli* biomass and thus be undetectable by direct PCR on bacteria in the water. Accordingly, we used overnight broth culture to generate a large quantity of *E. coli* cells, which were then subjected to analysis. We describe here the results of our surveillance of urban, suburban, and rural streams in the metropolitan Baltimore, Md., area for the prevalence of *tir*- and *stx*-positive *E. coli* from April 2002 to November 2004.

MATERIALS AND METHODS

Water sampling. The locations of the core group of nine sampling sites, most of which are located in the Gwynns Falls watershed, are shown in Fig. 1. These sites are sampled on a weekly basis 12 months of the year as part of the Baltimore Ecosystem Study, which is a component of the Long-Term Ecological Research Network (LTER) funded by the National Science Foundation. The Gwynns Falls watershed (76°30′, 39°15′) is ca. 163 km² and lies within the Piedmont Physiographic Province; the population in 2000 was ca. 356,000 people, with densities ranging from 2.2 to 19.4 persons per hectare (7). Pond Branch (POBR) is a stream located in a 100% forested small watershed, located in a suburban park (Oregon Ridge). Baisman's Run (BARN) is another small stream in a locale that is 63% forested, with low-density residential housing on septic tanks in some headwaters. Both of these streams lie in the Gunpowder watershed, which supplies drinking water to the city of Baltimore via Loch Raven Reservoir north of the city.

Gwynns Falls Gwynnbrook (GFGB) is a small stream associated with a 65% residential and 13% forested watershed. McDonough (MCDN) is a small stream located in a watershed associated with row crop (corn [*Zea mays*] and soybean [*Glycine max*]) agriculture; Gwynns Falls at Villa Nova (GFVN) is a medium-
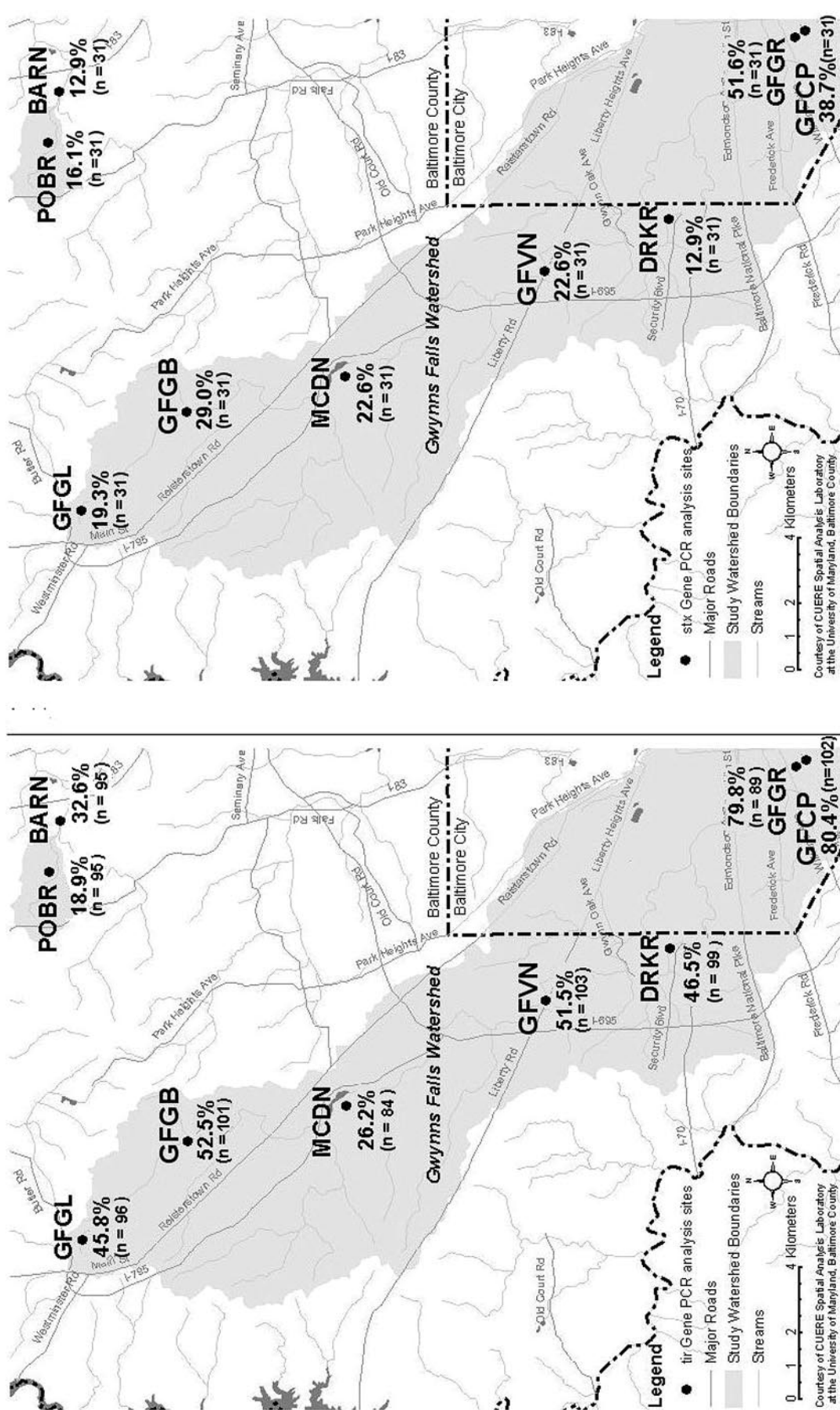
FIG. 1. Map depicting the core stream sampling sites in the Gwynns Falls watershed (~163 km^2) in Western Baltimore County/Baltimore City. Identities of site acronyms are given in the Materials and Methods. (Left) Number of samples assayed over the period from April 2002 to March 2004 and percentage of samples determined to be positive by *E. coli tir* gene PCR; (right) percentage of samples (n = 31) collected from July 2003 to March 2004 testing positive for either *E. coli stx*$_1$ or *stx*$_2$ genes.

sized stream that represents a transition between older and newer suburban housing; and Dead Run (DRKR), a medium stream, receives runoff from an older, suburban/commercial watershed. Gwynns Falls Gwynns Run (GFGR) represents a stream receiving substantial runoff from high-density urban housing, commercial, and industrial sources; >99% of the stream is subsurface, and the above-ground portion is ~100 m in length. GFGR flows into the Gwynns Falls <100 m upstream of the collection site at Carroll Park (GFCP). The Gwynns Falls in turn empties into the Patapsco River and then the Chesapeake Bay.

In addition to the nine core sites, other streams from the metropolitan Baltimore, MD area were sampled and assayed; those whose PCR amplicons were sequenced are listed as follows. In the Gwynns Falls watershed area, Gwynns Falls Glyndon–Main (GFGLM) and Gwynns Falls upstream of Gwynns Run (GFUGR) represent streams located in medium-density suburbs. Three sites from the Cub Hill neighborhood of northeastern Baltimore City/Baltimore County were also sampled weekly from 2003 onwards: these were included Jennifer Bridge Northwind Road, Jennifer Bridge Harford Hills, and Jennifer Bridge near Ontario street (JBON). Other sites sampled less regularly were Herring Run (HR), Jones Falls (JF), and Maiden's Choice (MC); these are urban streams, whereas Western Run (WR) and Spring Branch (SB) represent agricultural and suburban locales, respectively. Sites from Montgomery County, MD, included an urban stream (Rock Creek) close to the campus of the Walter Reed Army Institute for Research (WAL) and Near (railroad) Trestle (NT). Samples were also obtained from the Potomac River (POT), Fairfax County, Virginia (FFC), and from Little Cove Creek (CM) in southern Pennsylvania, located near Cove Mountain Farm, a dairy cattle operation that uses rotational grazing. One sample (MM) was obtained from Magothy Manor, a beachfront on the Magothy River in Annapolis, MD. During 2002 the nine core sites in the Gwynns Falls watershed were sampled several times a month (often weekly) from April through October, with sampling occurring once or twice a month during the colder months. The sampling protocol for these nine sites was increased to weekly in 2003 and 2004.

Water samples (500 ml) were collected from the sampling sites; delivered to the Beltsville Agricultural Research Center in Beltsville, MD; and subjected to culture (below) within 24 h of collection (with the exception of 5 weekly samples that were subjected to broth culture within 36 h of collection). Turbidity was measured by using a 20-ml aliquot on an Orbeco-Hellige turbidometer (Orbeco Analytical Systems, Inc., Farmingdale, NY). A total of 100 ml of water was filtered through a nitrocellulose, gridded, 47-mm-diameter, 0.45-μm-pore-size filter (catalog no. E04WG047S4; Osmonics, Westborough, MA) by using a Millipore (Bedford, MA) ground-glass filtration apparatus. The Millipore apparatus was rinsed with 70% ethanol between filtrations, and the plastic bottles used to collect the water samples were autoclaved for 20 min at 250°C between uses to prevent carryover of bacterial organisms from one collection to the next.

Filters were then placed in ~10 ml of minimal lactose broth (23) and cultured overnight at 37°C. The next morning several 1-ml aliquots were removed and subjected to centrifugation at 16,168 × *g* (13,000 rpm) for several minutes to pellet the bacterial cells. All but ~30 μl of the supernatant was withdrawn, and the pellet was subjected to DNA extraction by using Instagene matrix (8) and stored at 4° until assayed or at −20°C for long-term storage.

Coliform counts were obtained at the Baltimore Ecosystem Study laboratory at the University of Maryland, Baltimore County, by plating 1 ml of stream water (or serial dilutions, when the sample was anticipated to have a high concentration of microorganisms) on a Coliscan Easygel agar plate (Micrology Laboratories LLC, Goshen, IN), followed by incubation at 35°C for ~20 h. The numbers of colonies with a coloration (dark blue or purple) suggestive of *E. coli* were determined, and the values were multiplied by 100 to give counts per 100 ml.

***E. coli* isolates.** *E. coli* O157:H7 Odwalla strain was kindly provided by Pina Fratamico of the U.S. Department of Agriculture-Agricultural Research Service (USDA-ARS) Eastern Regional Research Laboratory, Wyndmoor, PA. Other *E. coli* strains were from a collection assembled by Marion Wachtel, Beltsville Agricultural Research Center, Beltsville, MD. Stream water-derived isolates of *E. coli* were obtained by using immunomagnetic separation reagents and a protocol previously described (23). Briefly, 10 ml of selected MLB-enriched water samples was incubated for 1 h with 40 ml of streptavidin beads (2.8 μm; Dynalbeads M-280 [Dynal A.S., Oslo, Norway] complexed to monoclonal anti-*E. coli* O157 antibodies; BioDesign International, Saco, ME) in a 1-ml volume of PBS-2 (150 mM potassium phosphate buffer [pH 7.2] and 150 mM NaCl). This incubation mix contained ca. 10^7 bacterial cells and 10^7 beads. After the 1 h of incubation, beads were shaken in PBS-2 for 15 min to dislodge poorly binding cells, condensed via magnetized holder, resuspended in 1 ml of PBS, and 100 ml plated onto MacConkey agar plates (two replicate plates) with an Autoplate 4000 spiral plater (Spiral Biotech, Norwood, MA). Plates were incubated overnight at 37°C, and individual colonies transferred to fresh MLB and replated on

MacConkey agar plates to confirm phenotype, after which a second round of immunomagnetic electrochemiluminescence (IM-ECL) screening was done on 20 colonies selected at random from the >100 colonies on the plate; colonies yielding strong positive signals were assayed by using the BBL Enterotube II (Becton Dickinson, Sparks, MD) and also subjected to DNA extraction and PCR with *lacZ* primers (see below) to confirm their identity as *E. coli*.

PCR and sequencing. Published TaqMan probes and primers were used to screen samples for the presence of the stx_1 and stx_2 genes via real-time multiplex PCR on a Stratagene Mx4000 instrument (22). The "Big Tir" primer set was used to amplify ~479-bp fragment of the *E. coli tir* gene via conventional PCR; the sequence of the forward primer was 5′-GGC GAC GGG TAT TGT ACA GG (nucleotides 738 to 757), and the sequence of the reverse primer was GCA GGC TTA TTC TCT ACC GTA CG (nucleotides 1216 to 1238 [5]). For amplification of the entire *tir* gene, we used the MS108+ and MS201− primer set described by DeVinney et al. (5); some isolates did not amplify with this primer set, so we substituted the forward primer Ntir (5′-ATG CCT ATT GGT AAY CTT GGY) for primer MS108+. Thermal cycling for *tir* gene detection was done in 50-μl volumes containing 1 U of *Taq* polymerase (Invitrogen, Gaithersburg, MD), 200 μM concentrations of each deoxynucleoside triphosphate, 1.5 mM MgCl$_2$, 5 μl of 10× PCR buffer, and 50 pmol of each primer. Cycling parameters for the ~479-bp region of the *tir* gene PCR were as follows: 95°C for 2 min, followed by 35 to 40 cycles of 95°C for 15 s, 60°C for 30 s, and 72°C for 1 min. Cycling conditions for the amplification of the entire *tir* gene were as follows: 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, 55°C for 1.5 min, and 72°C for 2 min. Dye terminator DNA sequencing of PCR amplicons was performed by using a BigDye v3.0 kit and an ABI 3100 model automated fluorescence sequencing instrument. Some amplicons (including those comprising the entire *tir* gene) were also cloned into *E. coli* by using the pGEM-T Vector System II (Promega, Madison, WI) and TOPO TA (Invitrogen, Gaithersburg, MD) kits and sequenced by using M13 forward and reverse primers. Sequence data were analyzed by using the Lasergene software package (DNASTar, Madison, WI). Phylogenetic trees were constructed from CLUSTAL W alignments by using neighbor-joining analysis with the Poisson correction (amino acid), complete deletion of positions containing gaps, and interior branch tests (1,000 bootstrap replications, 64238 random seed), rooted on *Citrobacter rodentium*, and using MEGA 2.1 (13).

For multidimensional scaling (MDS) analysis of phylogenetic data generated from the complete amino acid sequence alignments of the *tir* gene, the BioNumerics 2.5 software package was used (Applied Maths, Kortrijk, Belgium). Most commercial software packages for phylogenetic analyses do not offer this feature, mainly because the programming requires in-depth knowledge of matrix algebra and three-dimensional modeling; thus, a brief explanation may be in order. MDS is often used to interpret degrees of relatedness between observations; examples of published applications include microbial ecology, such as determining the indices of microbial diversity in soil (4); in bioinformatics, where it has been used to identify similar protein sequence motifs (17); and even in forensic science, where it has been used to model the propensity of serial killers to dispose of their victim's corpses in particular locales (15).

In the analysis presented here, taxa are represented as spheres located in a three-dimensional space, as opposed to the two-dimensional "tree" diagrams traditionally used in phylogenetic analyses. Consequently, relationships between taxa can be more easily visualized, since the graph containing the MDS arrangement can be rotated in around the *x*, *y*, or *z* axes and the exterior view can be "zoomed" in or out in order to more closely examine the proximity of one taxa to another. Taxa with >90% sequence homology may be represented in MDS as partially fused spheres rather than by adjacent branches, as in two-dimensional tree diagrams; taxa with 100% identical sequences can be represented by a single sphere with differently colored halves.

Alignments were generated by using GenBank depositions for a number of clinical isolates of *E. coli*: strains 95SF2 (*E. coli* O157:H−), 95ZG1 (*E. coli* O26:H−), EPEC 87A (*E. coli* O111), and 95NR1 (*E. coli* O111:H−) (accession nos. AF070067, AF070068, AF070069, and AF025311, respectively [20]), as well as *E. coli* O157:H7 strain 86-24 (accession no. AF125993 [5]) and strain B10, an *E. coli* O103:H2 serotype isolated from a rabbit (accession no. AF113597 [16]).

To determine the specificity of the Big Tir PCR primers, two panels of DNAs extracted from isolates of various pathogenic *E. coli* were assayed: the first panel was obtained from the collection of M. Wachtel, USDA, Beltsville, MD, and included *E. coli* O157 (*n* = 35), *E. coli* O103 (*n* = 2), and *E. coli* O111 (*n* = 3, nos. 83-1121, 82-0549, and 98-1208 of the *E. coli* Reference Center, University Park, PA); *E. coli* O26 (*n* = 2); and one each of *E. coli* O48, O91, O139, and O137 for a total of 46 samples. All isolates originated from clinical specimens, save for the *E. coli* O139 isolate, which were obtained from a pig with edema disease, and the O111 isolates 83-1121 and 98-1208 originating from a cow and
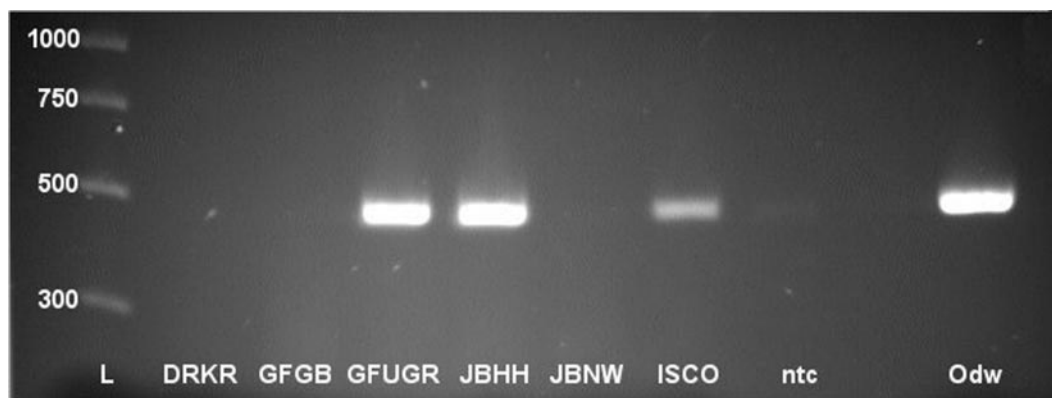
FIG. 2. Results of *E. coli tir* gene PCR conducted on coliforms cultured from stream water samples taken on 27 April 2004. Each lane represents a given stream site: DRKR, GFGB, Gwynns Falls up from Gwynns Run (GFuGR), Jennifer Bridge Harford Hills (JBHH), Jennifer Bridge Northwind Rd. (JBNW), and the Isco sampler from Rognel Heights (ISCO). Lanes: L, DNA ladder; ntc, no template control; lane Odw, *E. coli* O157:H7 Odwalla strain positive control.

a goat, respectively. The second panel was obtained from a collection of clinical isolates maintained by Choong Park at Fairfax Inova Hospital, Fairfax, VA, and consisted of *E. coli* O157 ($n = 4$), *E. coli* non-O157 ($n = 3$), *E. coli* O111 ($n = 4$), *E. coli* O45:H2 ($n = 3$), *E. coli* O26:H11 ($n = 4$) and one each of *E. coli* O76:H19, *E. coli* O145:NM, *E. coli* O103, and *E. coli* O103:H2, for a total of 22 samples. The quality of the isolate DNA was confirmed by performing PCR for the *lacZ* gene (10).

In addition to these isolates of *E. coli*, we also tested the *tir* primer set against the a panel of 18 bacteria isolated from the GFGR, MCDN, and POBR stream sites: *Enterobacter amnigenus*, *Enterobacter asburiae* ($n = 5$ isolates), *Enterobacter cancerogenus* ($n = 2$), *Citrobacter freundii* ($n = 2$), *Pantoea agglomerans*, *Klebsiella planticola*, *Klebsiella pneumoniae*, *Salmonella enterica* serovar Typhimurium, *Shigella flexneri*, *Shigella sonnei*, *Yersinia enterocolitica*, and *Serratia rubidaea*. These bacteria were identified via amplification of the 16S rRNA gene, sequencing of ca. 1000 to 1100 nucleotides of the gene, and similarity matrix querying of the Ribosomal Database Project II website (rdp.cme.msu.edu/html/). The queried sequence was required to have >97% similarity to a given database entry in order to be assigned a species identity (24). Also assayed were four isolates from the American Type Culture Collection (ATCC; Manassas, VA): *S. flexneri* (ATCC 12022, serotype 2b); *S. sonnei* (ATCC 25931); *Y. enterocolitica* (ATCC 23715), and *Prevotella loescheii* (ATCC 15930).

Statistical analyses. Proportions of total samples testing positive by *tir* and stx_1/stx_2 PCRs, for each of the nine stream sites, were analyzed via chi-square test by using the GraphPad Prism (GraphPad Software, San Diego, CA) and S-Plus (Mathsoft, Cambridge, MA) packages. In addition, the number of samples subjected to *tir* PCR testing was large enough to allow for multiple comparison testing between proportions by using an adaptation of the Dunnett test. For this purpose, *tir* PCR positive proportions (p) for each of the nine sites were transformed by using the arcsin transformation (p'). The POBR site was designated the control, and the other eight sites were compared to it by using the Dunnett test statistic q, with $v = 8$ and $k = 9$, i.e., $q_{0.05(2)}$ (29).

Nucleotide sequence accession numbers. The nucleotide sequences for the entire *tir* gene from the *E. coli* strains described in this study have been deposited with GenBank under the following accession numbers: SB isolate, AY944735; GFGB isolate, AY944736; BARN isolate, AY944737; and isolate no. 83-1121 (*E. coli* O111 from *Bos taurus*), AY944738.

RESULTS

Specificity of *tir* primers. The Big Tir primer set successfully amplified an ~479-bp fragment of this gene from all 39 isolates of *E. coli* O157 assayed, 3 of the 7 isolates of *E. coli* O111, 2 of the 6 isolates of *E. coli* O26, and 1 of the 4 isolates of *E. coli* O103. No amplicons were observed for the *E. coli* O139, *E. coli* O48, *E. coli* O137, *E. coli* O91, *E. coli* O145:NM, *E. coli* O76:H19, *E. coli* O45:H2, and non-O157 isolates. All of these *tir* PCR-negative samples successfully amplified with *lacZ*

primers, indicating that they were true negatives. DNA extracted from *C. freundii* did not amplify with either the Big Tir or Ntir/MS201− primer sets. None of the 18 non-*E. coli* bacterial isolates from stream water, as well as the four isolates obtained from the ATCC, amplified with the Big Tir primer set.

PCR screening of stream water bacterial cultures. As of 27 April 2004, 653 of 1,218 samples (53%) were determined to be positive by *tir* gene PCR. A representative PCR gel is shown in Fig. 2; in general, the use of an annealing temperature of 60°C resulted in clearly defined bands of ~500 bp. Positivity rates for the weekly *tir* PCR assays over the period from 25 April 2002 to 27 April 2004 are depicted in Fig. 3. A general observation is that positivity rates were highest in the summer months and lower in the winter months. For example, the data indicate that *tir*-positive coliforms were plentiful from April 2002 to November 2002, a time period that coincided with a sustained drought in the metropolitan Baltimore area. However, positivity rates were also high for the period from May 2003 to August 2003, a period which had substantially greater rainfall than the preceding year. These data suggest that, at least for the time period in which our observations were made, the presence of pathogenic *E. coli* in stream waters is not influenced by rainfall.

When data from the nine core sites were examined for the prevalence of *tir*-positive samples with data collected from April 2002 through March 2004, there were significant differences between sites (chi-square value of 137, 8 df, $P < 0.0001$). Interestingly, POBR and BARN, located in 100 and 63% forested watersheds, respectively, had unexpectedly high positivity rates of 18.9 and 2.6% (Fig. 1A). The suburban sites GFGL, GFGB, and GFVN had prevalence rates of 45.8, 52.5, and 51.5%, respectively, whereas MCDN, associated with row crop agriculture, had a prevalence rate of only 26.2%. Dead Run (DRKR) and GFGR, located in an urban area, had rates of 46.5 and 79.8%, respectively. Gwynns Falls Carroll Park, which receives flow from GFGR, had the highest prevalence of *tir* positivity at 80.4%. A more in-depth multiple comparison test between sites, made by using an adaptation of the Dunnett test, confirmed that there were significant differences between
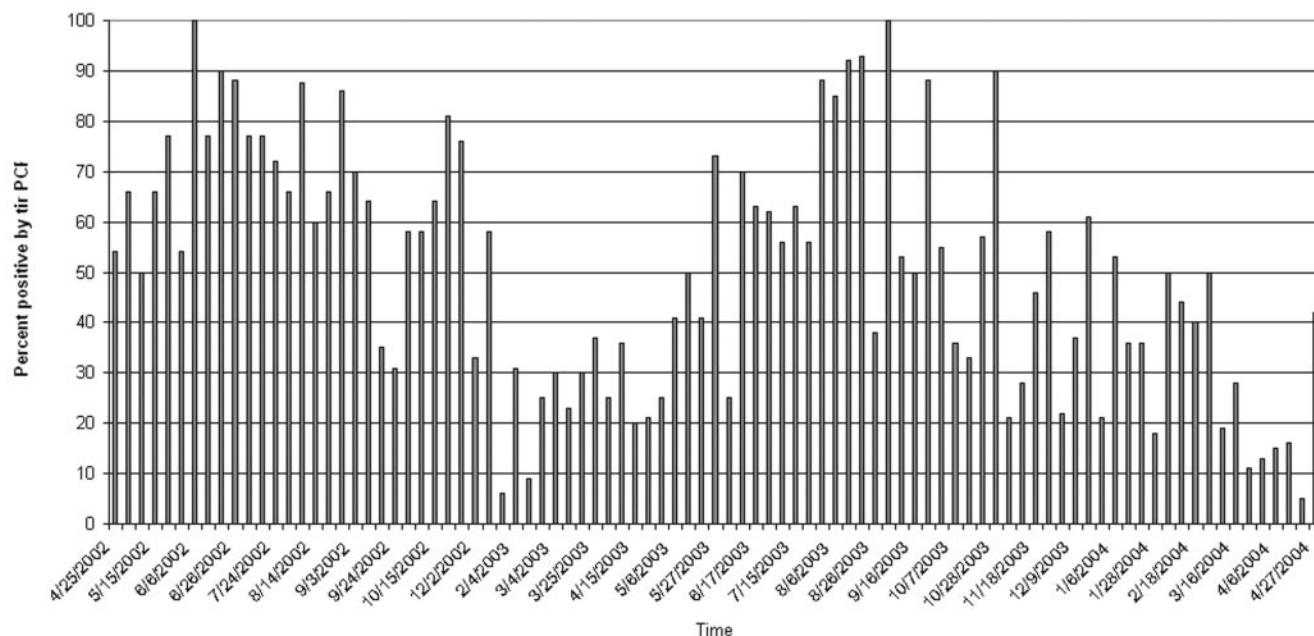
FIG. 3. Bar graph depicting percentage of weekly stream water-derived coliforms testing positive via *tir* gene PCR (*y* axis) during the period from 25 April 2002 to 27 April 2004 (*x* axis). The total number of samples was 1,218, with 653 (53%) positive. The mean number of samples per week was 12.7 (range, 9 to 21, *n* = 91 weeks).

POBR and GFGL (q' = 4.10 > $q_{0.05(2)8,9}$ = 3.35), GFGB (q' = 5.13), GFVN (q' = 4.87), DRKR (q' = 4.27), GFGR (q' = 8.85), and GFCP (q' = 9.24). There were no significant differences between POBR and BARN (q' = 2.31) and POBR and MCDN (q' = 1.21).
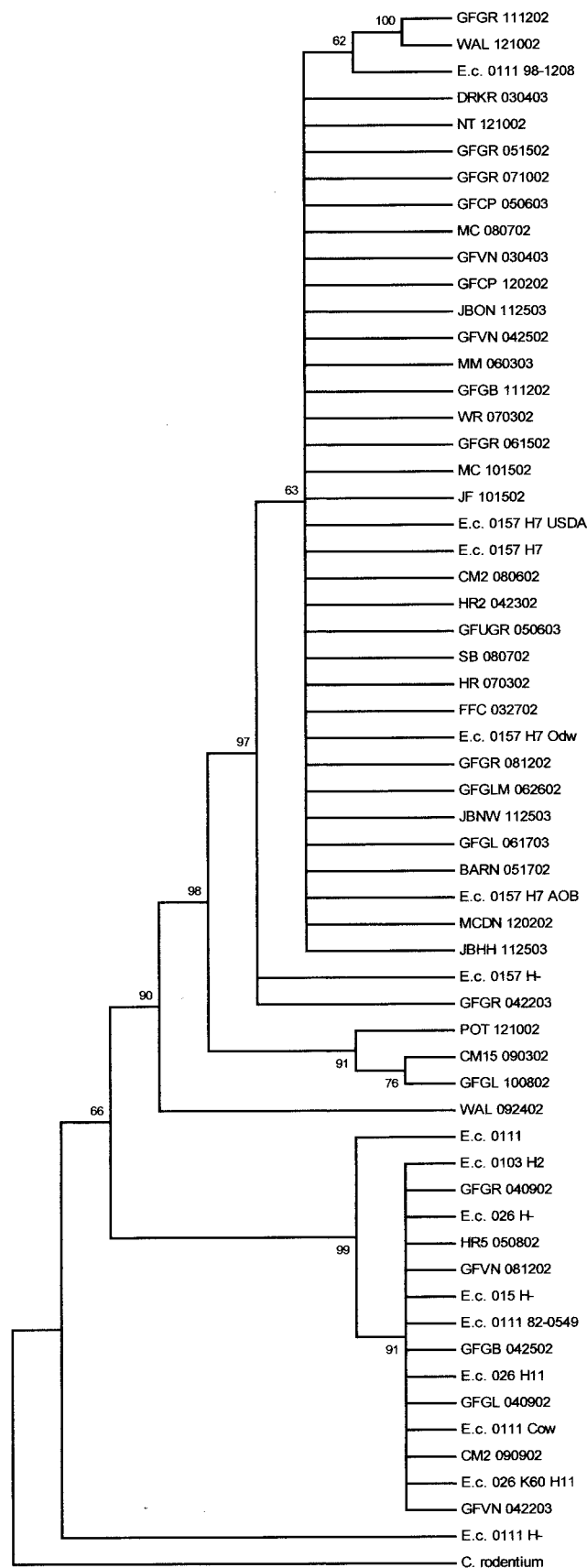
Coliform cultures from these nine sites, collected weekly from July 2003 to March 2004, were also assayed for the presence of stx_1 and stx_2, genes associated with EHEC, via multiplex real-time PCR. The data regarding the percentages of these samples (*n* = 31 each site) testing positive for either of these genes are presented in Fig. 1B. There were statistically significant differences between the nine sites in terms of the proportion of samples testing positive for one or both of the *stx* genes (chi-square value 21.9, 8 df, *P* = 0.005). As with the *tir* gene, we observed the highest prevalence of *stx* positivity in the more polluted urban streams, such as GFGR (51.6%) and GFCP (38.7%). Surprisingly, we did observe positive results from the two most pristine sites, POBR and BARN, neither of which are thought to be exposed to sewage or manure.

There were substantial differences between water samples collected from the urban and forested sites with regard to suspended solid content; for example, for 54 successive weekly collections made between March, 2003 and April 2004, the mean turbidity of the GFGR water was 32.89 ± 42.21 nephelometric turbidity units. For 56 successive collections made between February 2003 and April 2004, the mean turbidity for the POBR water was 3.04 ± 3.79 nephelometric turbidity units. Coliform counts between urban and forested sites also exhibited a marked difference; for successive collections taken from the period from 26 April 2002 to 6 May 2003, the mean count for GFGR was >20,000 per 100 ml (*n* = 38), for the agricul-

tural site MCDN the mean count was 521 per 100 ml (*n* = 33), and for POBR the mean count was 131 per 100 ml (*n* = 38).

Sequencing of the *tir* amplicons from coliform bacterial cultures and *E. coli* isolates. We sequenced *tir* amplicons (ca. 479 to 500 bp) amplified from 40 stream water bacterial enrichment cultures, as well as several clinical isolates of *E. coli* from our own collection, including *E. coli* O26:K60:H11, *E. coli* 026:H11, three strains of *E. coli* O111, and two strains of *E. coli* O157. When subjected to phylogenetic analysis, the translated amino acid sequences resolve into two major clades (Fig. 4). One clade contained a majority of the samples, which showed high degree (>99%) of sequence homology to the *tir* sequence of *E. coli* O157:H7 and *E. coli* strain 95SF2/*E. coli* O157:H− (with the exception of samples from GFGR from November 2002 and WAL from December 2002, which, surprisingly, had a greater homology with *E. coli* O111 strain 98-1208 from goats). The other clade contained a minority of sequences, with relatedness to those of clinical isolate 95ZG1/*E. coli* O26:H−, *E. coli* O26: H11, rabbit enterohemorrhagic strain B10/*E. coli* O103:H2, and a cow-derived *E. coli* O111 isolate.

The possibility exists that our stream water-derived coliform cultures may simultaneously contain multiple clones and/or strains of *tir*-bearing *E. coli* and, consequently, our sequences may represent chimeric assemblies from these simultaneously present clones. Therefore, attempts were made to obtain individual EHEC and/or EPEC *E. coli* cells by using the IM-ECL method, with the rationale that the anti-*E. coli* O157 immunoglobulin G would capture individual cells from the high background (10^7 to 10^8 cells per ml) of non-O157 *E. coli* cells present in the enriched cultures. These selected cells (recovered in volumes of 50 μl) were incubated on MacConkey agar

plates and, from among the >100 colonies present on these plates, 20 were randomly selected for another round of IM-ECL screening to confirm their identity as *E. coli* O157 cells, as well as ancillary testing with the BBL Enterotube II and *lac Z* PCR. A total of nine colonies from as many stream sites were ultimately chosen as representing likely *E. coli* O157 isolates, and three of these isolates had their entire *tir* gene amplified, cloned, and sequenced. We sequenced, in addition to these stream water isolates, the *tir* gene from *E. coli* O157:H7 Odwalla and from the *E. coli* O111 cow-derived isolate 83-1121.

An MDS plot, generated from phylogenetic analysis derived from comparison of the *tir* translated amino acid sequences of these isolates, is presented in Fig. 5. As expected, the Odwalla strain sequence showed a high degree of homology (>99%) with that of *E. coli* 86-24/O157:H7; the most heterogeneous interpretation of the nucleotide sequence of the Odwalla strain resulted in the substitution of two amino acids compared to strain 86-24. The September 2002 Gwynns Falls Glyndon (GFGL) and August 2002 Spring Branch (SB) isolates fell into a clade with relatedness to clinical isolate *E. coli* O111 (strain 87A). Interestingly, the cow-derived *E. coli* O111 isolate 83-1121 clusters with clinical isolate *E. coli* O26:H− and the rabbit EHEC strain O103:H2 (Fig. 4). We noted that the translated amino acid sequences of the latter two isolates are 100% identical and that the O26:H− isolate has only three nucleotide substitutions, out of 1,617, compared to isolate O103:H2. Given that these two isolates are from different genera of hosts, in different geographic locales, this finding was unexpected.
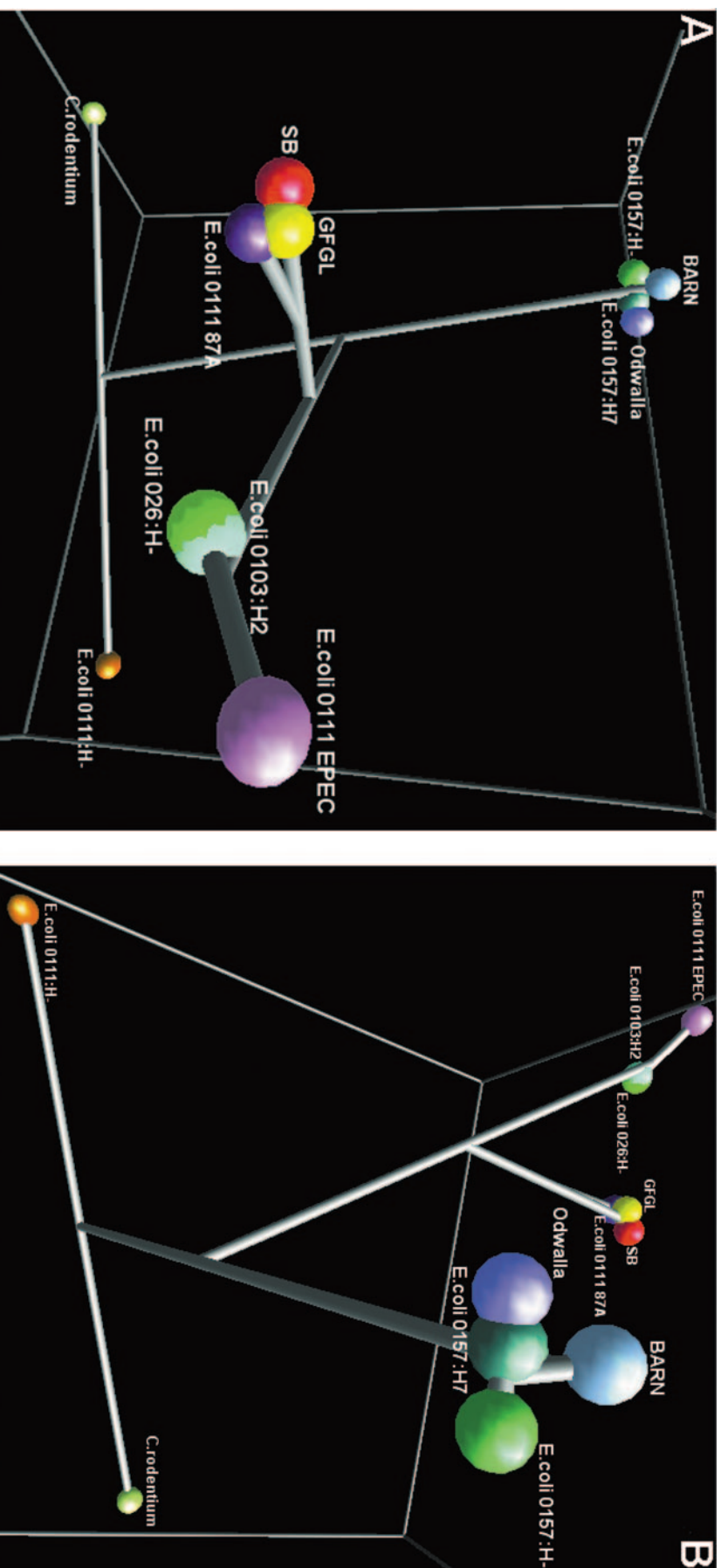
In our analyses the April 2002 BARN isolate tended to cluster with *E. coli* O157:H7 and *E. coli* O155:H-; however, the translated amino acid sequence of this isolate's *tir* gene revealed a number of interesting alterations compared to *E. coli* O157:H7. The BARN isolate had a 12-amino-acid insertion (GAGESKGAGESK) between amino acids 199 to 210 of *E. coli* O157:H7 and another 6-amino-acid insertion (TSTSTS) at amino acids 236 to 241 (Fig. 6). These insertions are not located in the putative transmembrane regions of the Tir protein (5). Serotype analysis of the three stream water isolates, conducted at the *E. coli* Reference Center in University Park, PA, indicated that SB and GFGL are *E. coli* O157, whereas the BARN isolate was *E. coli* O:63. None of these isolates tested positive with our *stx* multiplex PCR.

DISCUSSION

Our rationale for attempting to detect *tir*- and *stx*-containing strains of *E. coli* in stream water coliforms was that such strains

FIG. 4. *Citrobacter rodentium*-rooted neighbor-joining tree constructed from the alignment of a translated, 140-amino-acid region of the *tir* gene amplified from 40 stream water-derived coliform cultures. The identities of the stream sampling sites and USDA and GenBank reference sequences (i.e., taxa labeled as E.c. and C. rodentium) are given in Materials and Methods. Numbers associated with branches represent the percentage of bootstrapped replicates supporting that branching pattern. Numbers after stream samples represent the time (MM/DD/YY) the sample was collected; six-digit numbers after select *E. coli* O111 strains are reference numbers of the *E. coli* Reference Center.
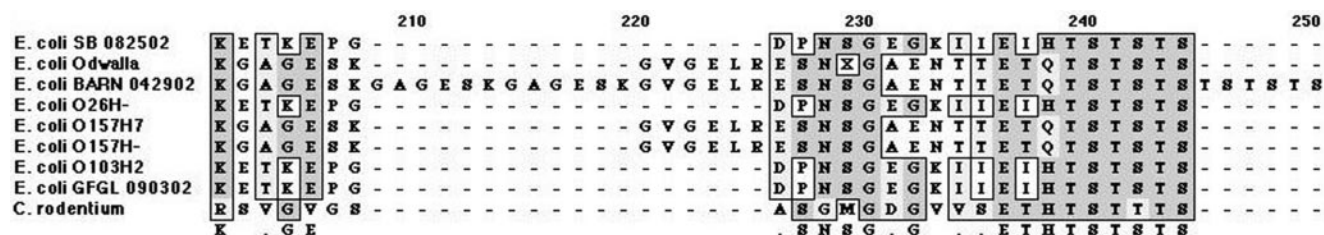
FIG. 5. MDS plot constructed from the phylogeny inferred from the alignment of the translated amino acid sequences (~582 amino acids in length) for the entire *tir* gene for three stream water samples (GFGL, SB, and BARN), *E. coli* O157:H7 Odwalla and *E. coli* O111 cow strains, and various *E. coli* reference sequences from GenBank (described in the Materials and Methods). Panels A and B represent different three-dimensional perspectives of the same dendrogram. Taxa are represented by colored spheres; the proximity of these spheres to one another reflects degrees of relatedness derived from amino acid sequence similarities. Note that the *E. coli* O26:H− and O103:H2 amino acid sequences are 100% identical; thus, their spheres are fused into one bicolored sphere.

FIG. 6. CLUSTAL W-generated alignment of translated amino acid sequences for the *tir* gene of *E. coli* showing insertions unique to the BARN sequence at positions 208 to 219 and positions 245 to 248 (equivalent to amino acids 199 to 210 and amino acids 236 to 241 of *E. coli* O157:H7 strain 86-24, accession no. AF125993). *E. coli* SB, Spring Branch isolate; *E. coli* GFGL, Gwynns Falls Glyndon isolate. Descriptions of accompanying reference serotypes and strains of *E. coli* are provided in Materials and Methods.

best represent bacterial populations associated with colonization of human and animal intestinal tracts and thus fecal contamination of the water source. Our studies indicate that *tir*- and *stx*-positive strains of *E. coli* are highly prevalent in polluted, urban stream waters. Although this finding is not unexpected, we were surprised to see evidence of *tir*- and *stx*-positive *E. coli* in streams located in forested watersheds, with POBR, in particular, receiving no input from water or waste generated by anthropogenic sources. Indeed, we suspect that a variety of potentially pathogenic *E. coli* strains may be present in our surface water samples. For example, when we performed a PCR assays for the EAST1 gene (28), an enterotoxin widely distributed among diarrheogenic *E. coli*, on coliforms cultured from 37 water samples obtained in July, 2003, all samples were positive.

The *E. coli* we detected are in a culturable metabolic state, which either means that they were recently deposited from human and/or animal feces or that actively metabolizing and reproducing populations of pathogenic *E. coli* are present in the stream water. Although there is convincing evidence that *E. coli* can maintain populations in tropical and subtropical freshwaters due to the combination of high temperatures and the availability of nutrients, the ability of this organism to persistently colonize freshwater habitat in temperate regions is unclear (27). It could be argued that in the summer months of temperate regions, some bodies of freshwater may attain temperature and nutrient levels similar to those of tropical regions and thus support reproducing populations of pathogenic *E. coli*. More detailed studies are necessary before either, or both, of these explanations can be validated. However, the presence of consistently high coliform and fecal coliform counts observed in samples from GFGR, as well as cysts of *Giardia intestinalis* (J. A. Higgins, unpublished data), would seem to indicate that discharges from leaking septic tanks and overwhelmed sewage distribution systems play a major role in the frequent deposition of feculent material into urban streams. Identifying sources of diarrheogenic *E. coli* in the rural and suburban streams of our study is more problematic. Bird species with an affinity for residing in urbanized areas, rats, and domestic animals such as dogs and cats may represent possible sources of *tir*- and *stx*-positive *E. coli* strains in the water samples subjected to testing. In addition, nonpoint source pollution from leaking septic tanks and other wastewater-containing facilities may contribute to waterborne *E. coli* populations.

The influence of rainfall and runoff events on the composition and concentration of fecal-associated bacteria has been the subject of numerous published studies; see, for example, the studies by Schiff and Kinney (21) and Kistemann et al. (12). Because we cannot quantitate EPEC or EHEC per se, we must rely on the total coliform count data in formulating conclusions about relationships between flow, rainfall, water temperature, and turbidity and the presence or concentration of these categories of *E. coli*. In general, these coliform count data show a fair degree of similarity to our prevalence data, i.e., lower counts or prevalence in Winter months. However, more detailed analyses of our sizeable hydrologic and meteorologic databases are necessary before any firm conclusions regarding the influence of these factors on the prevalence and distribution of the *tir*- and *stx*-positive *E. coli* in our sampling sites can be drawn.

The public health implications of our findings are uncertain. Analysis of the translated amino acid sequences for the ~479-bp *tir* fragment amplified from 40 stream sites and the entire ~1,677-bp *tir* gene from isolates from three stream sites indicates that the majority of our positives display a high degree of homology to *tir* sequences from EHEC O157:H7. However, it is unclear whether a high degree of sequence homology for given loci necessarily means that the *E. coli* present in the water are EPEC and/or EHEC; resolution of this question may require cell culture or animal model assays, which are beyond the scope of the work described here. If there are indeed EPEC and EHEC strains in the water, they appear to be present at low concentrations; by using an indirect quantitation method involving real-time PCR assays for the *lac Z*, *tir*, *stx*, and *eae* genes, conducted on DNA extracted from coliforms cultured from stream water samples, we have observed that the *tir*-, *eae*-, and *stx*-positive *E. coli* usually represent <1% of the overall (*lacZ*-positive) *E. coli* population in the sample (J. S. Karns, unpublished data). There is documented evidence that hemorrhagic colitis caused by *E. coli* can be acquired via swimming in freshwater (1) and, while we have witnessed people wading in and fishing from the streams in our survey, the depth of the water is not sufficient to allow for swimming. Thus, incidental contact with stream water may not constitute a significant health risk.

We are not aware of any publicized outbreaks of diarrheagenic *E. coli* among Baltimore City or Baltimore County residents during our stream surveillance. Is it possible that *tir* and *stx* positive *E. coli* are continually acquired and shed among healthy adults and children residing in our study area? A published report described an individual who excreted *stx*-positive *E. coli* rough:H21 strain cells for 5 months following diarrheal

disease (19), and some patients in a German study excreted *E. coli* O157 cells up to 124 days postinfection (9), but shedding in asymptomatic individuals has not been well characterized. Historically, in developed countries *tir*-containing strains of *E. coli*, such as EPEC, have been associated with outbreaks of diarrhea in infants, more rarely in adults (18). In their review, Levine and Edelman (14) cite studies on the carriage of EPEC in healthy children (<2 years old) in which as many as 20% of the subjects shed the organism in their feces. In a prospective study of 2,225 stool specimens originating from 445 Seattle children, 3.6% of the specimens were found to be positive for noncytotoxic, *eae*+, locally adherent *E. coli*, a prevalence rate greater than that of *E. coli* O157:H7, *Yersinia* spp., *Campylobacter* spp., *Salmonella* spp., and *Shigella* spp. (3). Obviously, further investigation is warranted, but our data suggest that *tir*- and *stx*-positive *E. coli* strains may be prevalent in metropolitan populations, perhaps as causes of otherwise unremarkable and underreported diarrheal illnesses.

REFERENCES

1. **Ackman, D., S. Marks, P. Mack, M. Caldwell, T. Root, and G. Birkhead.** 1997. Swimming-associated hemorrhagic colitis due to *Escherichia coli* O157:H7 infection: evidence of prolonged contamination of a freshwater lake. Epidemiol. Infect. **119:**1–8.
2. **Blank, T. E., J.-P. Nougayrede, and M. S. Donnenberg.** 2002. Enteropathogenic *Escherichia coli*, p. 81–118. *In* M. Donnenberg (ed.), *Escherichia coli*: virulence mechanisms of a versatile pathogen. Academic Press, Inc., New York, N.Y.
3. **Bokete, T. N., T. S. Whittam, R. A. Wilson, C. R. Clausen, C. M. O'Callahan, S. L. Moseley, T. R. Fritsche, and P. I. Tarr.** 1997. Genetic and phenotypic analysis of *Escherichia coli* with enteropathogenic characteristics isolated from Seattle children. J. Infect. Dis. **175:**1382–1389.
4. **Borneman, J., P. W. Skroch, K. M. O' Sullivan, J. A. Palus, N. G. Rumjanek, J. L. Jansen, J. Nienhuis, and E. W. Triplett.** 1996. Molecular microbial diversity of an agricultural soil in Wisconsin. Appl. Environ. Microbiol. **62:**1935–1943.
5. **DeVinney, R., M. Stein, D. Reinscheid, A. Abe, S. Ruschowski, and B. B. Finlay.** 1999. Enterohemorrhagic *Escherichia coli* O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated. Infect. Immun. **67:**2389–2398.
6. **DeVinney, R., J. L. Puente, A. Gauthier, D. Goosney, and B. B. Finlay.** 2001. Enterohaemorrhagic and enteropathogenic *Escherichia coli* use a different Tir-based mechanism for pedestal formation. Mol. Microbiol. **41:**1445–1458.
7. **Groffman, P. M., D. J. Blain, L. E. Band, K. T. Belt, G. S. Brush, J. M. Grove, R. V. Pouyat, I. C. Yesilonis, and W. C. Zipperer.** 2003. Down by the riverside: urban riparian ecology. Front. Ecol. Environ. **1:**315–321.
8. **Higgins, J. A., M. Jenkins, D. Shelton, R. Fayer, and J. Karns.** 2001. Rapid extraction of DNA from *Escherichia coli* and *Cryptosporidium parvum* for use in PCR. Appl. Environ. Microbiol. **67:**5321–5324.
9. **Karch, H., H. Russman, H. Schmidt, A. Schwarzkopf, and J. Heesemann.** 1995. Long-term shedding and clonal turnover of enterohemorrhagic *Escherichia coli* O157 in diarrheal diseases. J. Clin. Microbiol. **33:**1602–1605.
10. **Kator, H., and M. Rhodes.** 2003. Detection, enumeration, and identification of environmental organisms of public health significance, p. 113–144. *In* D. Moran and N. Horan (ed.), The handbook of water and wastewater microbiology Academic Press, Inc., New York, N.Y.
11. **Kenny, B., and B. Finlay.** 1997. Intimin-dependent binding of enteropathogenic *Escherichia coli* to host cells triggers novel signaling events, including tyrosine phosphorylation of phospholipase C-γ1. Infect. Immun. **65:**2528–2536.
12. **Kistemann, T., T. Claben, C. Koch, F. Dangendorf, R. Fischeder, J. Gebel, V. Vacata, and M. Exner.** 2002. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. Appl. Environ. Microbiol. **68:**2188–2197.
13. **Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei.** 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:**1244–1245.
14. **Levine, M. M., and R. Edelman.** 1984. Enteropathogenic *Escherichia coli* of classic serotypes associated with infant diarrhea: epidemiology and pathogenesis. Epidemiol. Rev. **6:**31–51.
15. **Lundrigan, S., and D. Canter.** 2001. A multivariate analysis of serial murderers' disposal site location choice. J. Environ. Psychol. **21:**423–432.
16. **Marches, O., J. P. Nougayrede, S. Boullier, J. Mainil, G. Charlier, I. Raymond, P. Pohl, M. Boury, J. De Rycke, E. Milon, and E. Oswald.** 2000. Role of Tir and intimin in the virulence of rabbit enteropathogenic *Escherichia coli* serotype O103:H2. Infect. Immun. **68:**2171–2182.
17. **Mendez, J., A. Falcon, and J. Lorenzo.** 2003. A procedure for biological sensitive pattern matching in protein sequences. Pattern Recognition Image Analysis Proc. **2652:**547–555.
18. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142–201.
19. **Orr, P., D. Milley, D. Colby, and M. Fast.** 1994. Prolonged fecal excretion of verotoxin-producing *Escherichia coli* following diarrheal illness. Clin. Infect. Dis. **19:**796–797.
20. **Paton, A. W., P. A. Manning, M. C. Woodrow, and J. C. Paton.** 1998. Translocated intimin receptors (Tir) of Shiga-toxigenic *Escherichia coli* isolates belonging to serogroups O26, O111, and O157 react with sera from patients with hemolytic-uremic syndrome and exhibit marked sequence heterogeneity. Infect. Immun. **66:**5580–5586.
21. **Schiff, K., and P. Kinney.** 2001. Tracking sources of bacterial contamination in stormwater discharges to Mission bay, California. Water Environ. Res. **73:**534–542.
22. **Sharma, V. K., E. A. Dean-Nystrom, and T. A. Casey.** 1999. Semi-automated fluorogenic PCR assays (TaqMan) for rapid detection of *Escherichia coli* O157:H7 and other Shiga toxigenic *E. coli*. Mol. Cell. Probes **13:**291–302.
23. **Shelton, D. R., and J. S. Karns.** 2001. Quantitative detection of *Escherichia coli* O157 in surface waters by using immunomagnetic electrochemiluminescence. Appl. Environ. Microbiol. **67:**2980–2985.
24. **Stackebrandt, E., and B. M. Goebel.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44:**846–849.
25. **Thorpe, C. M., J. M. Ritchie, and D. W. K. Acheson.** 2002. Enterohemorrhagic and other Shiga toxin-producing *Escherichia coli*, p. 119–154. *In* M. Donnenberg (ed.), *Escherichia coli*: virulence mechanisms of a versatile pathogen. Academic Press, Inc., New York, N.Y.
26. **Vallance, B. A., and B. B. Finlay.** 2000. Exploitation of host cells by enteropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA **16:**8799–8806.
27. **Winfield, M. D., and E. A. Groisman.** 2003. Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. Appl. Environ. Microbiol. **69:**3687–3694.
28. **Yamamoto, T., and P. Echeverria.** 1996. Detection of the enteroaggregative *Escherichia coli* heat-stable enterotoxin 1 gene sequences in enterotoxigenic *E. coli* strains pathogenic for humans. Infect. Immun. **64:**1441–1445.
29. **Zar, J.** 1999. Biostatistical analysis, 4th ed., p. 565. Prentice Hall, Inc., Englewood Cliffs, N.J.